RESEARCH   **Open Access**

CrossMark

Sequential Monte Carlo for inference of latent ARMA time-series with innovations correlated in time

Iñigo Urteaga[2], Mónica F. Bugallo[1*] and Petar M. Djurić[1]

Abstract

We consider the problem of sequential inference of latent time-series with innovations correlated in time and observed via nonlinear functions. We accommodate time-varying phenomena with diverse properties by means of a flexible mathematical representation of the data. We characterize statistically such time-series by a Bayesian analysis of their densities. The density that describes the transition of the state from time t to the next time instant $t + 1$ is used for implementation of novel sequential Monte Carlo (SMC) methods. We present a set of SMC methods for inference of latent ARMA time-series with innovations correlated in time for different assumptions in knowledge of parameters. The methods operate in a unified and consistent manner for data with diverse memory properties. We show the validity of the proposed approach by comprehensive simulations of the challenging stochastic volatility model.

Keywords: Sequential Monte Carlo, Correlated innovations, Latent time-series, State-space models, ARMA, FARIMA, Fractional Gaussian process

1 Introduction

This paper addresses inference of a broad class of latent time-series observed via nonlinear functions. We aim at modeling time-series with diverse memory properties in a unified manner so that a method for inference of heterogeneous time-varying data can be proposed. To that end, we elaborate on classical autoregressive moving average (i.e., ARMA) models and consider innovations[1] that are correlated in time. With these flexible modeling assumptions, a diverse set of scenarios and data properties can be accommodated. The studied latent time-series framework not only covers classical ARMA type models and their fractionally integrated generalizations, i.e., autoregressive fractionally integrated moving average (ARFIMA) processes but also allows for inference of time-series with heterogeneous memory properties.

The analysis of time-series is relevant in a plethora of disciplines in science, engineering and economics [1–3]. In all these areas, stochastic processes are used to model the behavior of time-varying data. Often, the modeling is

carried out by two processes, one of which is latent and the other, observed and informative about the hidden process.

Among the relevant features of time-series data and the stochastic models used for their description, their *memory* is one of the most important characteristics. On the one hand, there are short-memory processes, where only few past data values affect the present of the time-series. On the other, the present value is dependent on samples far into the past for long-memory processes.

ARMA models have been widely studied for characterizing short-term processes, as they accurately describe quickly forgetting data. The pioneering work on short-memory processes and ARMA(p, q) time-series was presented in the early 1950s by [4], it was continued by [5], and later expanded by [2]. ARMA(p, q) processes are defined by their autoregressive (AR) parameters $a_1, a_2, \cdots, a_p$, of order p; moving average (MA) parameters $b_1, b_2, \cdots, b_q$, of order q; and driving innovations u_t, which are assumed to be independent and identically distributed (i.i.d.).

The work on long-memory processes also began in the middle of the 20th century, with the groundwork laid by [6]. He studied Nile river data and realized that it manifested long-range dependence. In the following decades,

*Correspondence: monica.bugallo@stonybrook.edu
[1]Department of Electrical and Computer Engineering, Stony Brook University, 11794 Stony Brook, NY, USA
Full list of author information is available at the end of the article

Urteaga *et al. EURASIP Journal on Advances in Signal Processing* (2017) 2017:84

Page 2 of 15

plenty of other geophysical, climatological, and financial records have been described by similar long-term characteristics [7–9].

For modeling time-series with long memory, there are two types of formulations that have attracted interest of practitioners [8]. They arise naturally from limit theorems and classic models. With the first formulation, the long-memory processes are described as stationary increments of self-similar models, of which the fractional Gaussian process (fGp) is a prime example. The second formulation appears in the form of autoregressive fractionally integrated moving average processes. These models are built upon ARMA models by introducing non-integer values of the differencing parameter d, which accounts for the "integrated" part I of the model. The acronyms ARFIMA or FARIMA are used to refer to these processes (where the F refers to the "fractional" component), even if the ARIMA(p, d, q) notation suffices if fractional values of d are considered.

Both short- and long-memory processes (modeled by ARMA, FARIMA, or other models) are commonly used in practice to describe all kinds of time-varying data, including eolic phenomena [10, 11], biomedical signals [12], or financial markets [13, 14]. Our goal in this paper is to consider a generic formulation of time-series so that we accommodate diverse memory properties in a unified framework. We do so by considering ARMA models with innovations that are correlated in time.

We have pointed out that it is common to use two dependent processes to model observed phenomena. The reason is that often, we are not able to directly observe the process of interest. This may be due to the intricacies of the acquisition procedure or simply because the latent process cannot be observed. Inference of hidden processes is a very challenging task. The time-series estimation and prediction problems are certainly much more difficult when the process of interest is not directly observed, as is the case in this paper.

The latent process has dynamics that are not directly observed, but the observed process depends on the latent process in various forms. One represents the two by using a state-space formulation, where the evolution of the system is modeled over time by a series of hidden variables associated with another series of measurements. That is, the state-space model comprises a set of ordered observations y_t that depend on some latent time-evolving unknowns x_t. Oftentimes, one is interested in online or sequential estimation of the state, as data are observed over time.

The task of estimating the hidden processes (i.e., x_t) based on observations y_t has been widely studied, see [15] for example. To that end, two classes of problems are contemplated: one where the processes and the observations are linear functions of the states with additive

and Gaussian perturbations and another where the functions are nonlinear and/or the noises are not Gaussian. The former class allows for estimating the latent process by optimal methods (e.g., Kalman filtering [16]) while the latter, by resorting to suboptimal methods, based on Bayesian theory [17] or other approximating techniques [18]. Precisely, popular approaches are based on (1) model transformations (e.g., extended Kalman filtering [19]), (2) resorting to QML solutions [15], and (3) Monte Carlo sampling principles.

Our goal here is to provide a method that achieves sequential estimation of hidden time-series with diverse memory properties within a unified framework. We avoid Gaussianity and linearity assumptions in the model and thus consider techniques that can overcome such difficulties. In particular, among the advanced techniques available for online estimation of data, we choose to work with sequential Monte Carlo (SMC) or particle filtering (PF) methods [20–22].

Ever since the publication of [23], they have been shown to overcome the difficulties posed by non-Gaussian and nonlinear models. They have an extensive record of being successfully applied to many disciplines, including engineering [24], geophysical sciences [25], biology [26, 27], and economics [28]. Furthermore, it is an active area of research, where new paradigms and extensions to the classical SMC method for estimation of latent states and fixed parameters are being explored [29–31]. In this work, we focus on the use of standard SMC methods for inference of latent states with different memory properties in a unified and consistent manner. We present our contributions in detail after providing a brief synopsis of the paper in the following section.

1.1 Outline
We consider the general problem of sequential inference of latent time-series, observed via nonlinear functions. We describe the considered mathematical framework in Section 2, where we adopt a state-space representation of the processes. The latent time-series is modeled as an ARMA(p, q) process driven by innovations correlated in time. We make minimal assumptions about the observation equation so that any computable nonlinear function can be accommodated.

In Section 3, we provide a Bayesian analysis of the time-series model, for different assumptions about the parameters of the model. We first derive the joint probability density of the data, for which the computation of the covariance matrix is critical. We subsequently obtain the transition density of the time-series of interest, which plays a critical role in the proposed method.

We leverage the statistical description of the model and propose an SMC solution in Section 4. After an outline of the key concepts of the SMC technique in subsection 4.1,

Urteaga *et al. EURASIP Journal on Advances in Signal Processing* (2017) 2017:84

Page 3 of 15

in the following subsection, we present a novel SMC method for inference of latent time-series with non-i.i.d. innovations.

We conclude the paper with Section 5 where we evaluate the proposed set of SMC methods on an illustrative application: inference of the stochastic volatility of observed time-varying data. This model is not only of interest in practice (e.g., finance) but also serves as a challenging benchmark for alternative methods already available in the literature. The results validate our proposed method.

1.2 Contribution

The main contribution of this paper is a set of SMC methods for inference of latent states with different memory properties in a unified and consistent manner.

In previous work, SMC methods have been introduced for specific modeling assumptions: ARMA processes in [32–34] and latent fractional Gaussian processes in [35]. On the contrary, an SMC-based alternative for filtering time-series of different properties could be to pose the problem from a model selection perspective [36–38]. However, we study a different and more ambitious approach in this paper, as we target inference of latent states with different characteristics in a unified and consistent manner.

We consider time-varying phenomena with diverse characteristics, provide a flexible and compact mathematical representation of the data, and propose generic SMC methods for sequential inference of latent time-series with different memory properties.

More precisely:

- We provide a mathematical framework for the characterization of heterogeneous time-series as ARMA(p, q) models driven by innovations correlated in time.
- We derive a compact mathematical formulation that describes time-series of diverse memory properties.
- We derive both the joint and the transition density of such time-series under different model parameter assumptions.
- We present a set of SMC methods for inference of latent time-series with non-i.i.d. innovations, based on which parameters of the model are known.
- We demonstrate the performance of the proposed SMC on the stochastic volatility model driven by a fractional Gaussian process filtered by an ARMA(p, q) process.

2 Mathematical model

We are interested in the study of latent time-series with diverse memory properties observed via nonlinear functions. To systematically accommodate observed and hidden variables, we adopt the state-space methodology.

For the state, we use a flexible formulation that allows for short- and long-memory processes. As for the space equation, we make minimal assumptions.

Specifically, the state dynamics follow an ARMA(p, q) model with non-i.i.d. innovations. That is, the driving noise for the ARMA process is correlated in time. The only restriction is that the innovation process must be stationary. For the observation equation, we consider nonlinear functions of the state.

Mathematically, the model of interest is

$$\begin{cases} x_t = \sum_{i=1}^{p} a_i x_{t-i} + u_t + \sum_{j=1}^{q} b_j u_{t-j}\,, \\ y_t = h(x_t, v_t)\,, \end{cases} \tag{1}$$

where $x_t \in \mathbb{R}$ ($x_t = 0,\ \forall t < 0$) is the state of the latent process at t; $a_1, a_2, \cdots, a_p$ are the autoregressive (AR) parameters; and $b_1, b_2, \cdots, b_q$ are the moving average (MA) parameters of the ARMA model.

The symbol u_t represents a zero-mean Gaussian innovation process that is correlated in time. That is, it is fully characterized by its first (i.e., mean) and second moments (i.e., an arbitrary autocovariance function $\gamma_u(\tau)$). Note that stationarity is enforced (the mean is constant and the autocovariance depends only on the time lag τ), but no restriction on the form of the autocovariance function is imposed.

We denote with $y_t \in \mathbb{R}$ the observation at time t. We consider that the observation can be expressed by any generic function $h(x_t, v_t)$ and that the observation noise v_t can have any distribution for as long as the likelihood function $f(y_t|x_t)$ is computable up to a proportionality constant.

Our goal is to sequentially estimate the posterior distribution of x_t, given the observations $y_{1:t} \equiv \{y_1, y_2, \cdots, y_t\}$, that is, $f(x_t|y_{1:t})$. We want to update $f(x_t|y_{1:t})$ to $f(x_{t+1}|y_{1:t+1})$, for each new acquired observation y_{t+1}.

We resort to Bayesian theory for the derivation of such density. The key equation in the derivation is the one that connects the filtering density at t, $f(x_t|y_{1:t})$, with that at the next time instant $t + 1$, $f(x_{t+1}|y_{1:t+1})$. It is given by

$$f(x_{t+1}|y_{1:t+1}) = \int f(x_{t+1}|x_{1:t}, y_{1:t+1}) f(x_{1:t}|y_{1:t}) \mathrm{d}x_{1:t}. \tag{2}$$

The analytical solution to this integral is only possible for the case of Gaussian densities and linear functions, i.e., the celebrated Kalman Filter [16]. Since we do not restrict ourselves to such assumptions in this paper, we resort to SMC methods, widely popular since the seminal publication of [23]. These methods provide suboptimal solutions, as they compute probability random measure approximations of the densities of interest, while guaranteeing convergence under suitable conditions.

We detail in Section 3 the Bayesian analysis of the studied time-series (i.e., ARMA(p, q) models with correlated innovations), before delving into the intricacies of sequential Monte Carlo methods in Section 4.

3 Bayesian analysis of the time-series

We provide a Bayesian analysis of an ARMA(p, q) model driven by an innovation process that is not i.i.d. We emphasize the importance of considering the full ARMA(p, q) with correlated innovations model, as it allows for an accurate description of time-series with a wide range of memory properties within the same mathematical formulation.

Typical ARMA(p, q) models (i.e., with i.i.d. innovations) possess exponentially decaying memory properties. Regardless of the parameterization of the autoregressive or the moving average components, the time-series forgets its past at a rate proportional to $c^{-\tau}$ for a constant $c \in (0, 1)$. On the contrary, by considering correlated innovations u_t, we allow for the time-series to exhibit a wide range of decaying memory properties. That is, the modeled time-series can forget their past values sub-exponentially.

For example, if the innovations are fractional Gaussian processes, then the memory dependency is proportional to τ^{-c} for a constant $c \in (0, 1)$ (e.g., see Fig. 1). Note that such diverse memory characteristics cannot be modeled with ARMA models and uncorrelated innovations. Furthermore, any of the particular subcases (AR(p), MA(q), or ARMA(p, q), with or without correlated noise) are covered within the proposed generic formulation.

All in all, we study a generic ARMA(p, q) model with non-i.i.d. innovations. We are interested in the statistical properties of this stochastic process and, thus, in the joint distribution of the time-series up to time instant t, i.e.,
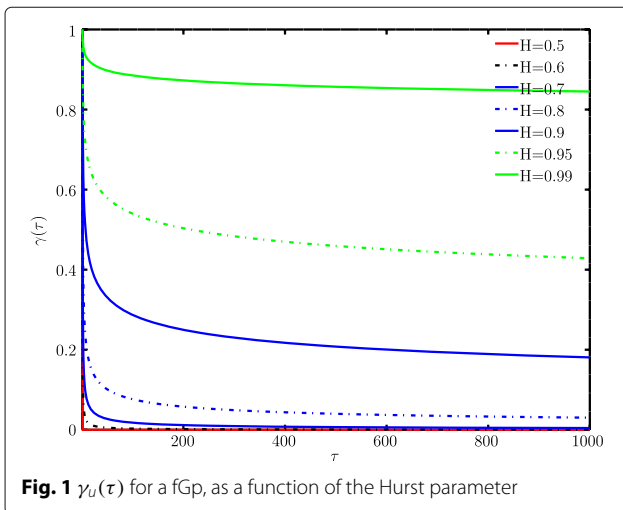

Fig. 1 $\gamma_u(\tau)$ for a fGp, as a function of the Hurst parameter

$f(x_{1:t})$. This distribution contains all the relevant information of the time-series, from which any marginal and conditional of interest can be derived.

We start by reformulating the state time-series in Eq. (1), where we fix $b_0 = 1$,

$$x_t = \sum_{i=1}^{p} a_i x_{t-i} + \sum_{j=0}^{q} b_j u_{t-j} \,,$$

$$1 - \sum_{i=1}^{p} a_i x_{t-i} = \sum_{j=0}^{q} b_j u_{t-j} \,, \tag{3}$$

and rewrite the above recursion in matrix form

$$
\begin{pmatrix}
1 & -a_1 & -a_2 & \cdots & -a_p & 0 & & 0 & 0 & \cdots & 0 \\
0 & 1 & -a_1 & \cdots & -a_{p-1} & -a_p & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & -a_{p-2} & -a_{p-1} & -a_p & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots & & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & & \cdots & 1 & -a_1 & -a_2 \\
0 & 0 & 0 & 0 & 0 & 0 & & 0 & \cdots & 1 & -a_1 \\
0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
x_t \\ x_{t-1} \\ x_{t-2} \\ \vdots \\ x_3 \\ x_2 \\ x_1
\end{pmatrix}
=
$$

$$
=
\begin{pmatrix}
1 & b_1 & b_2 & \cdots & b_q & 0 & & 0 & 0 & \cdots & 0 \\
0 & 1 & b_1 & \cdots & b_{q-1} & b_q & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & b_{q-2} & b_{q-1} & b_q & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots & & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & & \cdots & 1 & b_1 & b_2 \\
0 & 0 & 0 & 0 & 0 & 0 & & 0 & \cdots & 1 & b_1 \\
0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & \cdots & 1
\end{pmatrix}
\begin{pmatrix}
u_t \\ u_{t-1} \\ u_{t-2} \\ \vdots \\ u_3 \\ u_2 \\ u_1
\end{pmatrix}
\,. \tag{4}
$$

By defining the matrices

$$
A_t =
\begin{pmatrix}
1 & -a_1 & -a_2 & \cdots & -a_p & 0 & & 0 & 0 & \cdots & 0 \\
0 & 1 & -a_1 & \cdots & -a_{p-1} & -a_p & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & -a_{p-2} & -a_{p-1} & -a_p & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots & & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & & \cdots & 1 & -a_1 & -a_2 \\
0 & 0 & 0 & 0 & 0 & 0 & & 0 & \cdots & 1 & -a_1 \\
0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & \cdots & 1
\end{pmatrix}
\in \mathbb{R}^{t \times t},
\tag{5}
$$

and

$$
B_t =
\begin{pmatrix}
1 & b_1 & b_2 & \cdots & b_q & 0 & & 0 & 0 & \cdots & 0 \\
0 & 1 & b_1 & \cdots & b_{q-1} & b_q & 0 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & b_{q-2} & b_{q-1} & b_q & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots & & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & & \cdots & 1 & b_1 & b_2 \\
0 & 0 & 0 & 0 & 0 & 0 & & 0 & \cdots & 1 & b_1 \\
0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & \cdots & 1
\end{pmatrix}
\in \mathbb{R}^{t \times t},
\tag{6}
$$

we can immediately write $A_t x_{1:t} = B_t u_{1:t}$, and thus,

$$x_{1:t} = A_t^{-1} B_t u_{1:t}, \tag{7}$$

where the time subscript t also indicates matrix dimensionality.

In our problem of interest, the innovation process u_t is zero-mean Gaussian with an arbitrary autocovariance function $\gamma_u(\tau) = \sigma_u^2 \rho_u(\tau)$. The innovation process is correlated unless $\rho_u(\tau) = \delta(\tau)$, when we have i.i.d. noise. The vector of innovations up to time instant t, i.e., $u_{1:t}$, follows a zero-mean Gaussian multivariate density with covariance matrix $C_{u_t} = \sigma_u^2 R_{u_t}$, where

$$
R_{u_t} = \begin{pmatrix}
\rho_u(0) & \rho_u(1) & \cdots & \rho_u(t-2) & \rho_u(t-1) \\
\rho_u(1) & \rho_u(0) & \cdots & \rho_u(t-3) & \rho_u(t-2) \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\rho_u(t-2) & \rho_u(t-3) & \cdots & \rho_u(0) & \rho_u(1) \\
\rho_u(t-1) & \rho_u(t-2) & \cdots & \rho_u(1) & \rho_u(0)
\end{pmatrix}.
$$

(8)

Note the symmetric Toeplitz structure of the matrix, where the only required elements are the normalized autocovariance values $\rho_u(\tau)$, for lags $\tau = 0, 1, \cdots, t-1$.

Based on these sufficient statistics of the correlated innovation process and the formulation of the time-series as in Eq. (7), we derive the densities of interest for our model in Eq. (1).

3.1 Bayesian analysis: joint density of the time-series

Let us consider an ARMA(p, q) time-series with correlated innovations at time instant t, i.e.,

$$
x_{1:t} = A_t^{-1} B_t u_{1:t}, \qquad u_{1:t} \sim \mathcal{N}\left(0, \sigma_u^2 R_{u_t}\right).
$$

(9)

One readily concludes that the joint density of the time-series $x_{1:t}$ is a multivariate zero-mean Gaussian with a covariance matrix dependent on the matrices A_t, B_t, and R_{u_t}; that is,

$$
\begin{aligned}
x_{1:t} \sim f\left(x_{1:t}|\sigma_u^2\right) &= \mathcal{N}\left(x_{1:t}|0, \sigma_u^2 \Sigma_t\right), \\
\Sigma_t &= A_t^{-1} B_t R_{u_t} B_t^\top \left(A_t^{-1}\right)^\top \in \mathbb{R}^{t \times t}.
\end{aligned}
$$

(10)

It is also of practical interest to consider the unknown driving noise variance case. Instead of estimation of the unknown σ_u^2, we hereby proceed by marginalizing it, i.e., we Rao-Blackwellize σ_u^2. To that end, we use conjugate priors due to their convenient analytical properties on deriving the marginalized density.

We start with a scaled inverse chi-squared prior for the unknown σ_u^2,

$$
f\left(\sigma_u^2\right) = \chi^{-2}\left(\sigma_u^2|v_0, \sigma_0^2\right) = \frac{\left(\sigma_0^2 \frac{v_0}{2}\right)^{\frac{v_0}{2}}}{\Gamma\left(\frac{v_0}{2}\right)} \frac{1}{\left(\sigma_u^2\right)^{1+\frac{v_0}{2}}} e^{-\frac{v_0 \sigma_0^2}{2\sigma_u^2}}.
$$

(11)

The derivation of the marginalized joint density is then given by

$$
\begin{aligned}
f(x_{1:t}) &= \int_0^\infty f\left(x_{1:t}|\sigma_u^2\right) f\left(\sigma_u^2\right) d\sigma_u^2 \\
&= \int_0^\infty (2\pi)^{-\frac{t}{2}} \left(\sigma_u^2\right)^{-\frac{t}{2}} |\Sigma_t|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma_u^2} x_{1:t}^\top \Sigma_t^{-1} x_{1:t}} \\
&\quad \cdot \frac{\left(\sigma_0^2 \frac{v_0}{2}\right)^{\frac{v_0}{2}}}{\Gamma\left(\frac{v_0}{2}\right)} \frac{1}{\left(\sigma_u^2\right)^{1+\frac{v_0}{2}}} e^{-\frac{v_0 \sigma_0^2}{2\sigma_u^2}} d\sigma_u^2 \\
&\propto \int_0^\infty \left(\sigma_u^2\right)^{-\left(1+\frac{v_0+t}{2}\right)} e^{-\frac{1}{2\sigma_u^2}\left(v_0 \sigma_0^2 + x_{1:t}^\top \Sigma_t^{-1} x_{1:t}\right)} d\sigma_u^2 \\
&\propto \left[\left(v_0 \sigma_0^2 + x_{1:t}^\top \Sigma_t^{-1} x_{1:t}\right)\right]^{-\frac{v_0+t}{2}} \\
&\propto \left[\left(1 + \frac{1}{v_0} x_{1:t}^\top \left(\sigma_0^2 \Sigma_t\right)^{-1} x_{1:t}\right)\right]^{-\frac{v_0+t}{2}} \\
&= \mathcal{T}_{v_0}\left(x_{1:t}|0, \sigma_0^2 \Sigma_t\right).
\end{aligned}
$$

(12)

We conclude that the joint density of the time-series at time instant t after marginalization of the unknown variance is a multivariate Student t density

$$
\begin{aligned}
f(x_{1:t}) &= \mathcal{T}_v\left(x_{1:t}|\mu_t, \Phi_t\right) \\
&= \frac{\Gamma\left(\frac{v+t}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \pi^{\frac{t}{2}} v^{\frac{t}{2}} |\Phi_t|^{\frac{1}{2}}} \cdot \left(1 + \frac{(x_{1:t}-\mu_t)^\top \Phi_t^{-1}(x_{1:t}-\mu_t)}{v}\right)^{-\left(\frac{v+t}{2}\right)},
\end{aligned}
$$

(13)

where $v = v_0$ represents the degrees of freedom, $\mu_t = 0$ is the location parameter, and $\Phi_t = \sigma_0^2 \Sigma_t$ is the scale matrix.

3.2 Bayesian analysis: transition density of the time-series

In time-series analysis, the transition density of the process, which describes the dynamics of the state from time t to the next time instant $t+1$, is of practical interest. To that end, we leverage the joint densities derived in Subsection 3.1 for both the known and unknown innovation variance cases.

Let us consider the statistical description of the time-series at time instant $t+1$, i.e.,

$$
\begin{cases}
f\left(x_{1:t+1}|\sigma_u^2\right) = \mathcal{N}\left(x_{1:t+1}|0, \sigma_u^2 \Sigma_{t+1}\right), & \text{if } \sigma_u^2 \text{ is known,} \\
f(x_{1:t+1}) = \mathcal{T}_{v_0}\left(x_{1:t+1}|0, \sigma_0^2 \Sigma_{t+1}\right), & \text{if } \sigma_u^2 \text{ is unknown,}
\end{cases}
$$

(14)

and rewrite the covariance matrix Σ_{t+1} in block form

$$
\Sigma_{t+1} = \begin{pmatrix} h_{t+1} & \lambda_t \\ \lambda_t^\top & \Sigma_t \end{pmatrix}, \text{where} \begin{cases} h_{t+1} \in \mathbb{R}^{1\times 1} \\ \lambda_t \in \mathbb{R}^{1\times t} \\ \Sigma_t \in \mathbb{R}^{t\times t} \end{cases}.
$$

(15)

We can readily derive the transition density of the state by using the expressions for the conditionals of the multivariate Gaussian and Student t distributions [39]. Namely,

Urteaga *et al. EURASIP Journal on Advances in Signal Processing* (2017) 2017:84

Page 6 of 15

- if σ_u^2 is known

$$f\left(x_{t+1}|x_{1:t},\sigma_u^2\right) = \mathcal{N}\left(x_{t+1}|\mu_{t+1},\sigma_{t+1}^2\right),$$
$$\text{with} \begin{cases} \mu_{t+1} = \lambda_t \Sigma_t^{-1} x_{1:t} \,, \\ \sigma_{t+1}^2 = \sigma_u^2 \left(h_{t+1} - \lambda_t \Sigma_t^{-1} \lambda_t^\top\right) \,. \end{cases}$$
$$(16)$$

- if σ_u^2 is unknown

$$f(x_{t+1}|x_{1:t}) = \mathcal{T}_{\nu_{t+1}}\left(x_{t+1}|\mu_{t+1},\phi_{t+1}^2\right),$$
$$\text{with} \begin{cases} \nu_{t+1} = \nu_0 + t \,, \\ \mu_{t+1} = \lambda_t \Sigma_t^{-1} x_{1:t} \,, \\ \phi_{t+1}^2 = \frac{\nu_0\sigma_0^2 + x_{1:t}^\top \Sigma_t^{-1} x_{1:t}}{\nu_0 + t}\left(h_{t+1} - \lambda_t \Sigma_t^{-1} \lambda_t^\top\right) \,. \end{cases}$$
$$(17)$$

4 Proposed SMC method for inference of latent time-series

Our objective is to sequentially infer the evolution of a latent time-series with correlated innovations, as we acquire new observations. In Bayesian terminology, one is interested in updating the filtering density from $f(x_t|y_{1:t})$ to $f(x_{t+1}|y_{1:t+1})$ as new data are observed. We do so by using Eq. (2).

As previously pointed out, the analytical solution to such equation is intractable for the most interesting cases (models with nonlinearities and non-Gaussianities) and thus, we resort to sequential Monte Carlo methods. We briefly provide an overview of SMC methods in subsection 4.1, before explaining in detail our proposed method in subsection 4.2.

4.1 Sequential Monte Carlo

Monte Carlo methods are a class of computational algorithms that numerically approximate functions of interest by random sampling. In particular, sequential Monte Carlo methods recursively compute approximations to relevant probability densities, by replacing the true densities with discrete random probability measures

$$f(x) \approx f^M(x) = \sum_{m=1}^M w^{(m)} \delta\left(x - x^{(m)}\right),\qquad(18)$$

where $\delta(\cdot)$ is the Dirac delta function.

The points $x^{(m)}$ represent the support of the random measure and are called particles. These particles are assigned weights $w^{(m)}$, which are interpreted as probability masses. The random measure is thus a weighted sum of M particles and their weights.

The key to SMC methods is the sequential computation of Eq. (2), which is done by updating the approximating random measures at time instant t to the next time instant $t+1$. Let $f^M(x_t)$ be the approximation of $f(x_t|y_{1:t})$ at time instant t. The update of $f^M(x_t)$ to $f^M(x_{t+1})$ is done in two steps.

First, one propagates the particles $x_t^{(m)}$ to $x_{t+1}^{(m)}$ via a so-called proposal density $\pi(\cdot)$,

$$x_{t+1}^{(m)} \sim \pi\left(x_{t+1}|x_{1:t}^{(m)},y_{1:t+1}\right),\qquad(19)$$

where one may use all or part of the available information (that is, the history of observations and previous states). Then, one computes the weights of each candidate sample $x_{t+1}^{(m)}$ according to

$$w_{t+1}^{(m)} \propto w_t^{(m)} \frac{f\left(y_{t+1}|x_{t+1}^{(m)}\right)f\left(x_{t+1}^{(m)}|x_{1:t}^{(m)}\right)}{\pi\left(x_{t+1}^{(m)}|x_{1:t}^{(m)},y_{1:t+1}\right)},\qquad(20)$$

where $f\left(y_{t+1}|x_{t+1}^{(m)}\right)$ is the likelihood of the new observation given sample $x_{t+1}^{(m)}$, and $f\left(x_{t+1}^{(m)}|x_{1:t}^{(m)}\right)$ is the transition density of the latent state. The computation of the weights is followed by their normalization so that they sum up to one and form a proper probability random measure.

SMC methods require an additional third step called resampling [40]. If one proceeds with propagation and weight computation only, the approximation $f^M(x_t)$ degenerates quickly, as only few of the particles are assigned non-negligible weights. Resampling consists on deciding which particles to propagate by selecting those with higher probability, i.e., bigger weights $w_t^{(m)}$. One prevents the quick deterioration of the SMC method by resorting to resampling methods (see [40] for an overview of the most common techniques). These are often triggered based on the effective sample size of the SMC approximation at every time instant [41].

The choice of the proposal density is critical for any SMC method. It has been shown that the optimal importance function is $f(x_{t+1}|x_t,y_{1:t+1})$, which minimizes the variance of the resulting random measure. However, this density is analytically intractable in our problem of interest. We adopt the simpler, but yet effective, alternative known as Sequential Importance Resampling (SIR) [23].

In summary, we sample new particle candidates from the transition density of the latent states $f(x_{t+1}|x_{1:t})$. Such proposal function entails that the weighting of the particles is proportional to their likelihood function, i.e., $w_{t+1}^{(m)} \propto w_t^{(m)} f\left(y_{t+1}|x_{t+1}^{(m)}\right)$. Details of the proposed SMC method for inference of latent ARMA models with correlated innovations follows.

4.2 Proposed SMC method

We first present an SMC method for inference of an ARMA process with correlated innovations, when the ARMA parameters, i.e., $\theta = (a_1 \; a_2 \; \cdots \; a_p \; b_1 \; b_2 \; \cdots \; b_q)^\top$, are known. We later relax the assumptions for the case when these parameters are unknown. In all the cases,

the normalized autocovariance values $\rho_u(\tau)$ for lags $\tau = 0, 1, \cdots, t-1$, of the correlated innovation process must be known.

4.2.1 Proposed SMC method: known ARMA parameters

Let us consider at time instant t the following probability random measure approximation of the filtering density $f(x_t | y_{1:t})$:

$$f^M(x_t) = \sum_{m=1}^{M} w_t^{(m)} \delta \left(x_t - x_t^{(m)} \right). \qquad (21)$$

Upon reception of a new observation y_{t+1}, the algorithm proceeds as follows:

1. Compute the joint normalized covariance matrix Σ_{t+1} at time instant $t+1$.

$$\Sigma_{t+1} = A_{t+1}^{-1} B_{t+1} R_{u_t} B_{t+1}^{\top} \left(A_{t_1}^{-1} \right)^{\top} = \begin{pmatrix} h_{t+1} & \lambda_t \\ \lambda_t^{\top} & \Sigma_t \end{pmatrix}. \qquad (22)$$

2. Perform resampling of the state's genealogical line by drawing from a categorical distribution defined by the random measure $f^M(x_t)$.

$$\bar{x}_{1:t}^{(m)} \sim \left\{ x_t^{(m)}, w_t^{(m)} \right\}, \text{where } m = 1, \cdots, M. \quad (23)$$

3. Propagate the state particles by sampling from the transition density, conditioned on the available resampled streams $\bar{x}_{1:t}^{(m)}$.

   - If σ_u^2 is known

$$x_{t+1}^{(m)} \sim \mathcal{N} \left(x_{t+1} | \mu_{t+1}^{(m)}, \sigma_{t+1}^2 \right),$$
$$\text{where } \begin{cases} \mu_{t+1}^{(m)} = \lambda_t \Sigma_t^{-1} \bar{x}_{1:t}^{(m)}, \\ \sigma_{t+1}^2 = \sigma_u^2 \left(h_{t+1} - \lambda_t \Sigma_t^{-1} \lambda_t^{\top} \right). \end{cases} \qquad (24)$$

   - If σ_u^2 is unknown

$$x_{t+1}^{(m)} \sim \mathcal{T}_{\nu_{t+1}} \left(x_{t+1} | \mu_{t+1}^{(m)}, \phi_{t+1}^{2(m)} \right),$$
$$\text{where } \begin{cases} \nu_{t+1} = \nu_0 + t, \\ \mu_{t+1}^{(m)} = \lambda_t \Sigma_t^{-1} \bar{x}_{1:t}^{(m)}, \\ \phi_{t+1}^{2(m)} = \frac{\nu_0 \sigma_0^2 + \bar{x}_{1:t}^{(m)\top} \Sigma_t^{-1} \bar{x}_{1:t}^{(m)}}{\nu_0 + t} \left(h_{t+1} - \lambda_t \Sigma_t^{-1} \lambda_t^{\top} \right). \end{cases} \qquad (25)$$

4. Compute the non-normalized weights for the drawn particles according to

$$\widetilde{w}_{t+1}^{(m)} \propto f \left(y_{t+1} | x_{t+1}^{(m)} \right), \qquad (26)$$

and normalize them to obtain a new random measure

$$f^M(x_{t+1}) = \sum_{m=1}^{M} w_{t+1}^{(m)} \delta \left(x_{t+1} - x_{t+1}^{(m)} \right). \qquad (27)$$

For the above method to be applicable, one needs to have full knowledge of the parameters in the transition density. That is, the matrices A_{t+1}, B_{t+1}, and $R_{u_{t+1}}$ must be known for the covariance matrix Σ_{t+1} to be computed for propagation of the state particles.

One can efficiently compute Σ_{t+1} by leveraging algebraic tricks prompted by the structural properties of the involved matrices (Toeplitz and upper triangular). On the one hand, the upper triangular nature of A_t and B_t simplify the number of computations considerably (the inverse of an upper triangular matrix is also upper triangular). The product $A_t^{-1} B_t$ is a matrix with a structure similar to A_t and B_t: an upper triangular matrix with elements of its first row shifted to the right. On the other hand, due to the Toeplitz structure of the R_{u_t} matrix, one can resort to a Levinson-Durbin type technique [42] to recursively compute the necessary matrix product operations.

The assumption that knowledge of the parameters within A_{t+1} and B_{t+1} exists, however, is often not substantiated. Therefore, we resort to a parameter sampling scheme when the ARMA parameters are not known. We augment the state vector with the unknown parameters $\rho_t = (x_t \, \theta_t)^{\top}$, similar to the work in [43, 44]. Note that the subscript t in θ_t does not imply that the parameter evolves over time. It is there only to signify that we obtain samples of the unknowns at time t.

The full parameter posterior for the model in Eq. (1) is analytically intractable, and thus, we cannot draw samples from the true parameter posterior. Furthermore, as the parameters do not change over time, their particle propagation becomes troublesome and various methodologies have been suggested to overcome these challenges. Some include the use of artificial parameter evolution [23], while others resort to kernel smoothing [43] or density-assisted (DA) particle filtering techniques [45].

In this paper, we explore and compare two sampling alternatives, one based on the principles of DA-SMC methods and another where importance sampling (IS) of the parameters is carried out. In the former, one approximates the posterior of the unknown parameters with a density of choice; in the latter, one draws from a proposal density for the parameters and later adjusts by computing the appropriate weights.

These proposed methods are the first approximation to dealing with unknown ARMA parameters. We acknowledge that any of the advanced SMC techniques that mitigate the challenges of estimating constant parameters (e.g., parameter smoothing [29, 46, 47] or nested SMC

Urteaga *et al. EURASIP Journal on Advances in Signal Processing* (2017) 2017:84

Page 8 of 15

methods [30, 31]) can only improve the accuracy of the proposed SMC methods.

4.2.2 Proposed SMC method: DA-SMC for unknown ARMA parameters

We now propose an SMC method for the case where the parameters of the latent ARMA model, i.e., $\theta = (a_1 \ a_2 \ \cdots \ a_p \ b_1 \ b_2 \ \cdots \ b_q)^\top$, are unknown. This first alternative follows the principles of density-assisted SMC methods. Because the true posterior of the unknown parameters is analytically intractable, it approximates such posterior with a density of choice.

In particular, we propose to approximate the posterior of the unknown parameter θ, given the current time-series $x_{1:t}$, with a Gaussian distribution, i.e.,

$$f\left(\theta_{t+1}^{(m)}|x_{1:t}^{(m)}\right) \approx \mathcal{N}\left(\theta_{t+1}|\mu_{\theta_t}, \Sigma_{\theta_t}\right), \tag{28}$$

where the sufficient statistics are computed based on samples and weights available at this time instant

$$\mu_{\theta_t} = \sum_{i=1}^M w_t^{(m)} \theta_t^{(m)},$$
$$\Sigma_{\theta_t} = \sum_{i=1}^M w_t^{(m)} \left(\theta_t^{(m)} - \mu_{\theta_t}\right)\left(\theta_t^{(m)} - \mu_{\theta_t}\right)^\top. \tag{29}$$

One uses this approximation to propagate parameter samples from this time instant to the next. As a result, the overall weight computation of the SMC method simplifies to

$$\widetilde{w}_{t+1}^{(m)} \propto f\left(y_{t+1}|x_{t+1}^{(m)}\right) \cdot \frac{f\left(x_{t+1}^{(m)}|x_{1:t}^{(m)},\theta_{t+1}^{(m)}\right)}{\pi(x_{t+1})} \cdot \frac{f\left(\theta_{t+1}^{(m)}|x_{1:t}^{(m)}\right)}{\pi(\theta_{t+1})}$$
$$\propto f\left(y_{t+1}|x_{t+1}^{(m)}\right). \tag{30}$$

In summary, the proposed DA-SMC for the unknown parameter case considers a joint state and parameter random measure at time instant t of the following form

$$f^M(\rho_t) = \sum_{m=1}^M w_t^{(m)} \delta\left(\rho_t - \rho_t^{(m)}\right), \tag{31}$$

and, upon reception of a new observation y_{t+1}, proceeds as follows:

1. Estimate the sample mean and covariance of the parameter vector θ_t.

$$\begin{cases} \mu_{\theta_t} = \sum_{i=1}^M \theta_t^{(m)} w_t^{(m)}, \\ \Sigma_{\theta_t} = \sum_{i=1}^M \left(\theta_t^{(m)} - \mu_{\theta_t}\right)\left(\theta_t^{(m)} - \mu_{\theta_t}\right)^\top w_t^{(m)}. \end{cases} \tag{32}$$

2. Draw new parameter samples from the Gaussian approximation to the posterior density with the newly computed sufficient statistics.

$$\theta_{t+1}^{(m)} \sim f\left(\theta_t|x_{1:t}^{(m)}\right) \approx \mathcal{N}\left(\theta_{t+1}|\mu_{\theta_t}, \Sigma_{\theta_t}\right). \tag{33}$$

3. Compute the joint covariance matrix for each parameter sample $\theta_{t+1}^{(m)}$.

$$\Sigma_{t+1}^{(m)} = A_{t+1}^{(m)^{-1}} B_{t+1}^{(m)} R_{u_t} B_{t+1}^{(m)^\top} A_{t_1}^{(m)^{-1^\top}} = \begin{pmatrix} h_{t+1}^{(m)} & \lambda_t^{(m)} \\ \lambda_t^{(m)^\top} & \Sigma_t^{(m)} \end{pmatrix}. \tag{34}$$

4. Perform resampling of the state's genealogical line by drawing from a categorical distribution defined by the random measure $f^M(x_t)$.

$$\overline{x}_{1:t}^{(m)} \sim \left\{x_t^{(m)}, w_t^{(m)}\right\}, \text{where } m = 1, \cdots, M. \tag{35}$$

5. Propagate the state particles by sampling from the transition density, conditioned on available resampled streams $\overline{x}_{1:t}^{(m)}$.

   - If σ_u^2 is known

$$x_{t+1}^{(m)} \sim \mathcal{N}\left(x_{t+1}|\mu_{t+1}^{(m)}, \sigma_{t+1}^{2(m)}\right),$$
$$\text{where} \begin{cases} \mu_{t+1}^{(m)} = \lambda_t^{(m)} \Sigma_t^{(m)^{-1}} \overline{x}_{1:t}^{(m)}, \\ \sigma_{t+1}^{2(m)} = \sigma_u^2 \left(h_{t+1}^{(m)} - \lambda_t^{(m)} \Sigma_t^{(m)^{-1}} \lambda_t^{(m)^\top}\right). \end{cases} \tag{36}$$

   - If σ_u^2 is unknown

$$x_{t+1}^{(m)} \sim \mathcal{T}_{v_{t+1}}\left(x_{t+1}|\mu_{t+1}^{(m)}, \phi_{t+1}^{2(m)}\right),$$
$$\text{where} \begin{cases} v_{t+1} = v_0 + t, \\ \mu_{t+1}^{(m)} = \lambda_t^{(m)} \Sigma_t^{(m)^{-1}} \overline{x}_{1:t}^{(m)}, \\ \phi_{t+1}^{2(m)} = \frac{v_0 \sigma_0^2 + \overline{x}_{1:t}^{(m)^\top} \Sigma_{1:t}^{(m)^{-1}} \overline{x}_{1:t}^{(m)}}{v_0 + t} \\ \quad \times \left(h_{t+1}^{(m)} - \lambda_t^{(m)} \Sigma_t^{(m)^{-1}} \lambda_t^{(m)^\top}\right). \end{cases} \tag{37}$$

6. Compute the non-normalized weights for the drawn particles according to

$$\widetilde{w}_{t+1}^{(m)} \propto f\left(y_{t+1}|x_{t+1}^{(m)}\right), \tag{38}$$

and normalize them to obtain a new probability random measure

$$f^M(\rho_{t+1}) = \sum_{m=1}^M w_{t+1}^{(m)} \delta\left(\rho_{t+1} - \rho_{t+1}^{(m)}\right). \tag{39}$$

4.2.3 Proposed SMC method: IS-SMC for unknown ARMA parameters

We now propose an alternative SMC method for the unknown ARMA parameter case too, based on importance sampling principles. Instead of approximating the analytically intractable parameter posterior, one can choose a proposal density and apply IS to jointly adjust the state and parameter samples.

Specifically, we use a Gaussian proposal density to draw samples for the unknown ARMA parameters θ. At every

Urteaga *et al. EURASIP Journal on Advances in Signal Processing* (2017) 2017:84

Page 9 of 15

time instant, one propagates parameter particles by sampling from the proposal

$$\pi(\theta_{t+1}) = \mathcal{N}\left(\theta_{t+1}|\mu_{\theta_t}, \Sigma_t\right), \tag{40}$$

with sufficient statistics as in Eq. (29). The corresponding weight computation results in

$$
\begin{aligned}
\widetilde{w}_{t+1}^{(m)} &\propto f\left(y_{t+1}|x_{t+1}^{(m)}\right) \cdot \frac{f\left(x_{t+1}^{(m)}|x_{1:t}^{(m)}\theta_{t+1}^{(m)}\right)}{\pi(x_{t+1})} \cdot \frac{f\left(\theta_{t+1}^{(m)}|x_{1:t}^{(m)}\right)}{\pi(\theta_{t+1})} \\
&= f\left(y_{t+1}|x_{t+1}^{(m)}\right) \cdot \frac{f\left(\theta_{t+1}^{(m)}|x_{1:t}^{(m)}\right)}{\mathcal{N}(\theta_{t+1}|\mu_{\theta_t},\Sigma_{\theta_t})} \ .
\end{aligned}
\tag{41}
$$

Since the posterior of the parameters is analytically intractable, we have

$$
f\left(\theta_{t+1}^{(m)}|x_{1:t}^{(m)}\right) = \frac{f\left(x_{1:t}^{(m)}|\theta_{t+1}^{(m)}\right)f\left(\theta_{t+1}^{(m)}\right)}{f\left(x_{1:t}^{(m)}\right)} \propto \frac{f\left(x_{1:t}^{(m)}|\theta_{t+1}^{(m)}\right)}{f\left(x_{1:t}^{(m)}\right)} ,
\tag{42}
$$

which results in

$$
\begin{cases}
f\left(\theta_{t+1}^{(m)}|x_{1:t}^{(m)}\right) \propto \dfrac{\mathcal{N}\left(x_{1:t}^{(m)}\big|0,\sigma_u^2\Sigma_t^{(m)}\right)}{\mathcal{N}\left(x_{1:t}^{(m)}\big|0,\sigma_u^2\Sigma_t^{(\mu_{\theta_t})}\right)}, & \text{if } \sigma_u^2 \text{ is known,} \\[3ex]
f\left(\theta_{t+1}^{(m)}|x_{1:t}^{(m)}\right) \propto \dfrac{\mathcal{T}_{v_0}\left(x_{1:t}^{(m)}\big|0,\sigma_0^2\Sigma_t^{(m)}\right)}{\mathcal{T}_{v_0}\left(x_{1:t}^{(m)}\big|v_0,0,\sigma_0^2\Sigma_t^{(\mu_{\theta_t})}\right)}, & \text{if } \sigma_u^2 \text{ is unknown.}
\end{cases}
\tag{43}
$$

With $\Sigma_t^{(\mu_{\theta_t})}$, we describe the covariance matrix computed using the parameter estimates μ_{θ_t} as in Eq. (29), while with $\Sigma_t^{(m)}$, we refer to the covariance matrix evaluated per drawn parameter sample $\theta_{t+1}^{(m)}$.

Therefore, the proposed IS-SMC for the unknown parameter case at time instant t starts with a joint state and parameter random measure

$$
f^M(\rho_t) = \sum_{m=1}^{M} w_t^{(m)} \delta\left(\rho_t - \rho_t^{(m)}\right), \tag{44}
$$

and, upon reception of a new observation y_{t+1}, proceeds as follows:

1. Estimate the sample mean and covariance of the parameter vector θ_t.

$$
\begin{aligned}
\mu_{\theta_t} &= \sum_{i=1}^{M} \theta_t^{(m)} w_t^{(m)}, \\
\Sigma_{\theta_t} &= \sum_{i=1}^{M} \left(\theta_t^{(m)} - \mu_{\theta_t}\right)\left(\theta_t^{(m)} - \mu_{\theta_t}\right)^{\top} w_t^{(m)} \ .
\end{aligned}
\tag{45}
$$

2. Draw new parameter samples from the Gaussian proposal with the newly computed sufficient statistics.

$$
\theta_{t+1}^{(m)} \sim \pi(\theta_{t+1}) = \mathcal{N}\left(\theta_{t+1}|\mu_{\theta_t}, \Sigma_{\theta_t}\right) \ . \tag{46}
$$

3. Compute the joint normalized covariance matrix for each parameter sample $\theta_{t+1}^{(m)}$.

$$
\Sigma_{t+1}^{(m)} = A_{t+1}^{(m)^{-1}} B_{t+1}^{(m)} R_{u_t} B_{t+1}^{(m)^{\top}} A_{t_1}^{(m)^{-1^{\top}}} = \begin{pmatrix} h_{t+1}^{(m)} & \lambda_t^{(m)} \\ \lambda_t^{(m)^{\top}} & \Sigma_t^{(m)} \end{pmatrix}.
\tag{47}
$$

4. Perform resampling of the state's genealogical line by drawing from a categorical distribution defined by the random measure $f^M(x_t)$.

$$
\bar{x}_{1:t}^{(m)} \sim \left\{ x_t^{(m)}, w_t^{(m)} \right\}, \text{where } m = 1, \cdots, M. \tag{48}
$$

5. Propagate the state particles by sampling from the transition density, conditioned on available resampled streams $\bar{x}_{1:t}^{(m)}$.

   - If σ_u^2 is known

$$
x_{t+1}^{(m)} \sim \mathcal{N}\left(x_{t+1}|\mu_{t+1}^{(m)}, \sigma_{t+1}^{2(m)}\right) ,
$$
$$
\text{where } \begin{cases} \mu_{t+1}^{(m)} = \lambda_t^{(m)} \Sigma_t^{(m)^{-1}} \bar{x}_{1:t}^{(m)}, \\ \sigma_{t+1}^{2(m)} = \sigma_u^2 \left(h_{t+1}^{(m)} - \lambda_t^{(m)} \Sigma_t^{(m)^{-1}} \lambda_t^{(m)^{\top}}\right) . \end{cases}
\tag{49}
$$

   - If σ_u^2 is unknown

$$
x_{t+1}^{(m)} \sim \mathcal{T}_{v_{t+1}}\left(x_{t+1}|\mu_{t+1}^{(m)}, \phi_{t+1}^{2(m)}\right) ,
$$
$$
\text{where } \begin{cases} v_{t+1} = v_0 + t, \\ \mu_{t+1}^{(m)} = \lambda_t^{(m)} \Sigma_t^{(m)^{-1}} \bar{x}_{1:t}^{(m)}, \\ \phi_{t+1}^{2(m)} = \frac{v_0\sigma_0^2 + \bar{x}_{1:t}^{(m)^{\top}} \Sigma_t^{(m)^{-1}} \bar{x}_{1:t}^{(m)}}{v_0 + t} \\ \qquad \times \left(h_{t+1}^{(m)} - \lambda_t^{(m)} \Sigma_t^{(m)^{-1}} \lambda_t^{(m)^{\top}}\right) . \end{cases}
\tag{50}
$$

6. Compute the non-normalized weights for the drawn particles.

   - If σ_u^2 is known

$$
\widetilde{w}_{t+1}^{(m)} \propto \frac{f\left(y_{t+1}|x_{t+1}^{(m)}\right) \cdot \mathcal{N}\left(x_{1:t}^{(m)}\Big|0, \sigma_u^2 \Sigma_t^{\left(\theta_{t+1}^{(m)}\right)}\right)}{\mathcal{N}\left(\theta_{t+1}^{(m)}|\mu_{\theta_t}, \Sigma_{\theta_t}\right)\mathcal{N}\left(x_{1:t}^{(m)}\Big|0, \sigma_u^2 \Sigma_t^{(\mu_{\theta_t})}\right)} .
\tag{51}
$$

   - If σ_u^2 is unknown

$$
\widetilde{w}_{t+1}^{(m)} \propto \frac{f\left(y_{t+1}|x_{t+1}^{(m)}\right) \cdot \mathcal{T}_{v_0}\left(x_{1:t}^{(m)}\Big|0, \sigma_0^2 \Sigma_t^{\left(\theta_{t+1}^{(m)}\right)}\right)}{\mathcal{N}\left(\theta_{t+1}^{(m)}|\mu_{\theta_t}, \Sigma_{\theta_t}\right)\mathcal{T}_{v_0}\left(x_{1:t}^{(m)}\Big|v_0, 0, \sigma_0^2 \Sigma_t^{(\mu_{\theta_t})}\right)} .
\tag{52}
$$

Urteaga *et al. EURASIP Journal on Advances in Signal Processing* (2017) 2017:84

Page 10 of 15

and normalize them to obtain a new probability random measure

$$f^M(\rho_{t+1}) = \sum_{m=1}^{M} w_{t+1}^{(m)} \delta\left(\rho_{t+1} - \rho_{t+1}^{(m)}\right). \quad (53)$$

5 Practical application

We now illustrate the applicability of the proposed SMC methods and evaluate their performance. We do so by considering the stochastic log-volatility (SV) state-space framework. That is, the observations are a zero-mean process with time-varying log-variance that one wants to estimate.

The SV model is popular in the study of nonlinear state-space models (due to the estimation challenges that it presents [15, 48–50]) and is of interest in finance (due to its applicability in the study of stock returns [17, 51–53]).

It has been established that Kalman filter (KF)-based methods fail to accurately estimate the latent state for SV models. In principle, for the nonlinearities in the SV model observation equation, extensions to the popular KF, such as the extended Kalman filter (EKF) [19], the
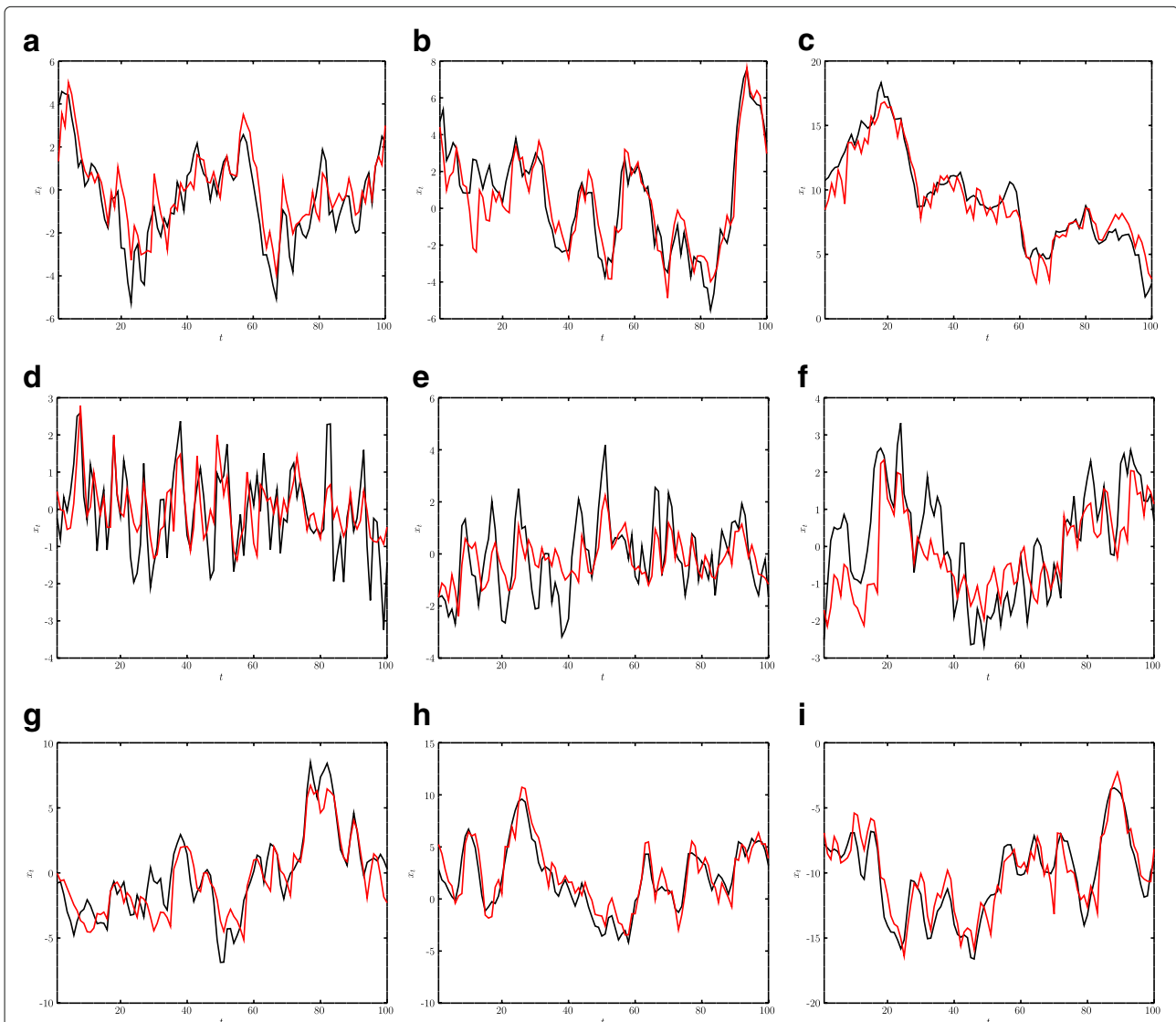
Fig. 2 True (black) and estimated state (red) for the proposed SMC method with known ARMA parameters and unknown σ_u^2 **a** AR(1), $a_1 = 0.85, H = 0.5$. **b** AR(1), $a_1 = 0.85, H = 0.7$. **c** AR(1), $a_1 = 0.85, H = 0.9$. **d** MA(1), $b_1 = 0.8, H = 0.5$. **e** MA(1), $b_1 = 0.8, H = 0.7$. **f** MA(1), $b_1 = 0.8, H = 0.9$. **g** ARMA(1, 1), $a_1 = 0.85, b_1 = 0.8, H = 0.5$. **h** ARMA(1, 1), $a_1 = 0.85, b_1 = 0.8, H = 0.7$. **i** ARMA(1, 1), $a_1 = 0.85, b_1 = 0.8, H = 0.9$.

Urteaga *et al. EURASIP Journal on Advances in Signal Processing* (2017) 2017:84

Page 11 of 15

Table 1 MSE performance of the proposed SMC methods for ARMA models (known a and b) with fGn, known and unknown σ_u^2

PF type	State estimation error (MSE)	
	Known σ_u^2	Unknown σ_u^2
AR(1), $H = 0.5$	1.1081	1.1945
AR(1), $H = 0.7$	1.3946	1.4397
AR(1), $H = 0.9$	1.1195	1.1970
MA(1), $H = 0.5$	1.0223	1.0686
MA(1), $H = 0.7$	1.0585	1.1136
MA(1), $H = 0.9$	0.87374	0.94053
ARMA(1,1), $H = 0.5$	1.5947	1.6197
ARMA(1,1), $H = 0.7$	1.7852	1.8516
ARMA(1,1), $H = 0.9$	1.7214	1.7362

unscented Kalman filter (UKF) [54], and other Sigma-Point Kalman filters [55] should be applicable. However, as reported in [56], these methods fail when addressing the SV model since they are unable to update their prior beliefs for such model (the Kalman gain is always null). Alternatives based on transformations of the model have been suggested ([15, 50]) but, as reported in [33], they fall short when compared to SMC methods.

Furthermore, the SV model has been in use in econometrics for a long time [53], as it is of interest in estimating the risk involved in financial transactions. There, the observations describe the price evolution of an asset, for which estimating its volatility is critical. This is not an easy task, and many efforts have been reported, where the risk is described with diverse memory properties [7, 14, 57].

Motivated by its practical application and the challenges it poses to the inference problem, we focus on the SV model, where the log-volatility is described by a latent ARMA(p, q) model with correlated innovations.

Without loss of generality, we focus on ARMA models with fractional Gaussian noise. With this modeling, we accommodate a wide range of memory properties: from uncorrelated to long-memory processes. This is a natural extension of the classical ARMA model, where instead of i.i.d. Gaussian innovations, the ARMA(p, q) filters a fractional Gaussian process with Hurst parameter H. The properties of such model are equivalent to those of the FARIMA(p, d, q), when $d = H - \frac{1}{2}$.

Mathematically, the SV model, where the latent time-series is an ARMA(p, q) with fractional Gaussian noise, is written as

$$\begin{cases} x_t = \sum_{i=1}^{p} a_i x_{t-i} + \sum_{j=1}^{q} b_j u_{t-j} + u_t, \\ y_t = e^{\frac{x_t}{2}} v_t, \end{cases} \tag{54}$$
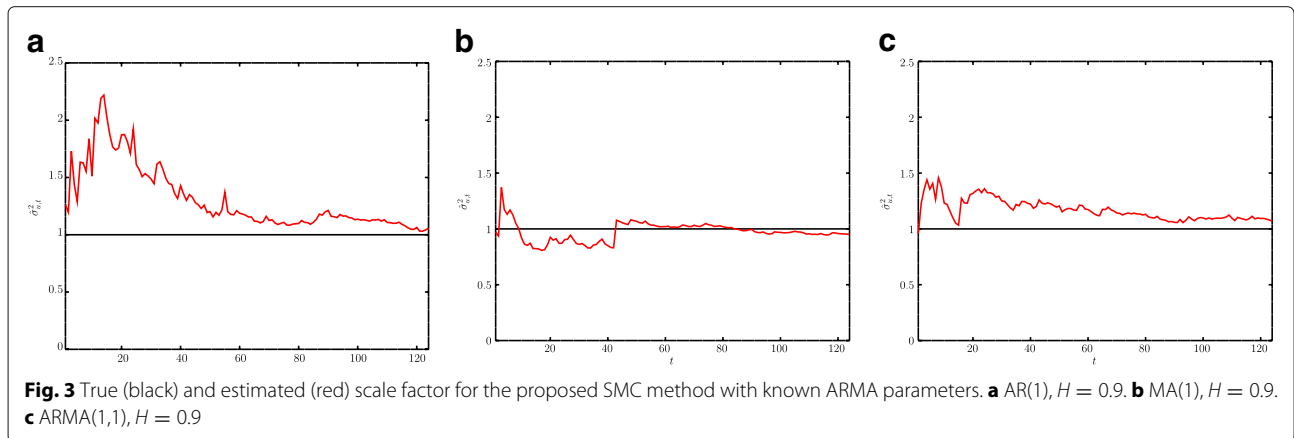
where v_t is a standard Gaussian variable and the state innovation u_t is a zero-mean Gaussian process with known autocovariance function $\gamma_u(\tau)$. In particular, for the fractional Gaussian process,

$$\gamma_u(\tau) = \frac{\sigma_u^2}{2} \left[|\tau - 1|^{2H} - 2 |\tau|^{2H} + |\tau + 1|^{2H} \right], \tag{55}$$

which is parameterized by the Hurst parameter H and variance σ_u^2. When $H = 0.5$, the process is uncorrelated, while the memory of the innovations increases as $H \to 1$. We illustrate in Fig. 1 the dependency of the normalized autocovariance function (i.e., $\sigma_u^2 = 1$) with respect to the parameter H.

We evaluated the proposed method in this nonlinear model, first under the known ARMA parameter case, and then, under unknown parameters. We show in Fig. 2 how the proposed method is able to accurately track different latent processes with diverse memory properties, even when the innovation variance σ_u^2 is unknown. The SMC methods were run with $M = 1000$ particles for different values of the Hurst parameter.

We further studied the filtering performance of the methods described in Subsection 4.2.1 and conclude, based on results summarized in Table 1, that the proposed SMC method successfully estimates the latent state, both for the known and unknown innovation variance cases, for any given memory.

Fig. 3 True (black) and estimated (red) scale factor for the proposed SMC method with known ARMA parameters. **a** AR(1), $H = 0.9$. **b** MA(1), $H = 0.9$. **c** ARMA(1,1), $H = 0.9$

Urteaga *et al. EURASIP Journal on Advances in Signal Processing* (2017) 2017:84

Page 12 of 15

Table 2 MSE performance of the proposed SMC methods for ARMA models (unknown a and b) with fGn, known and unknown σ_u^2

PF type	State estimation error (MSE)					
	Known a, b	Known a, b	Unknown a, b, DA	Unknown a, b, , IS	Unknown a, b, DA	Unknown a, b IS
	Known σ_u^2	Unknown σ_u^2	Known σ_u^2	Known σ_u^2	Unknown σ_u^2	Unknown σ_u^2
AR(1), $H = 0.5$	1.0991	1.127	1.6689	1.7337	1.4549	1.5903
AR(1), $H = 0.7$	1.4077	1.4375	2.5759	5.9889	1.9272	3.191
AR(1), $H = 0.9$	1.1336	1.1774	2.4334	6.5974	1.7795	6.4853
MA(1), $H = 0.5$	1.0348	1.0758	1.1033	1.185	1.1384	1.3831
MA(1), $H = 0.7$	1.0878	1.1138	1.1857	1.2688	1.1884	1.3748
MA(1), $H = 0.9$	0.88045	0.90841	0.96348	1.1124	0.97747	1.2517
ARMA(1,1), $H = 0.5$	1.638	1.6512	2.8563	3.6266	2.3157	2.3619
ARMA(1,1), $H = 0.7$	1.7452	1.7926	3.0939	4.1174	2.7807	2.4627
ARMA(1,1), $H = 0.9$	1.7374	1.7533	4.3466	20.617	2.5818	2.569

Note that the unknown σ_u^2 case induces a slight loss in accuracy. However, the estimation performance is comparable. The justification relies on the form of the derived marginalized density. As more data are observed, the transition density for the unknown variance case (i.e., Eq. (17)) becomes very similar to the one in the known case (i.e., Eq. (16)). This occurs because a Student t distribution with high degrees of freedom is very similar to a Gaussian distribution. Thus, the proposal densities in both SMC methods become almost identical with time.

To get further insight on the impact of not knowing the innovation variance σ_u^2, we study the evolution of the scale factor in Eq. (17) over time. We plot (see Fig. 3) the scale factor

$$\frac{v_0 \sigma_0^2 + x_{1:t}^\top \Sigma_t^{-1} x_{1:t}}{v_0 + t} \tag{56}$$

as we get more data and observe that the estimate approaches the true σ_u^2 value. The estimation accuracy improves with time for all the evaluated ARMA parameterizations and memory properties of the innovation process.

We now turn our attention to the more challenging scenario where the ARMA parameters are unknown. We evaluate both proposed approaches, i.e., the DA- and IS-based SMC methods from subsections 4.2.2 and 4.2.3, respectively.

Once again, we study the performance of the method for latent processes with different memory properties. In Table 2, we provide averaged state mean squared error (MSE) results for AR(1), MA(1), and ARMA(1,1) models with uncorrelated ($H = 0.5$), short- ($H = 0.7$), and long-memory ($H = 0.9$) fractional Gaussian innovations.

The state filtering results allow us to conclude that both proposed approaches are suitable solutions for inference of latent processes with unknown ARMA parameters. Besides, it is noticeable that the impact on state estimation accuracy of not knowing the parameters a and b is more pronounced than not knowing the innovation variance σ_u^2.
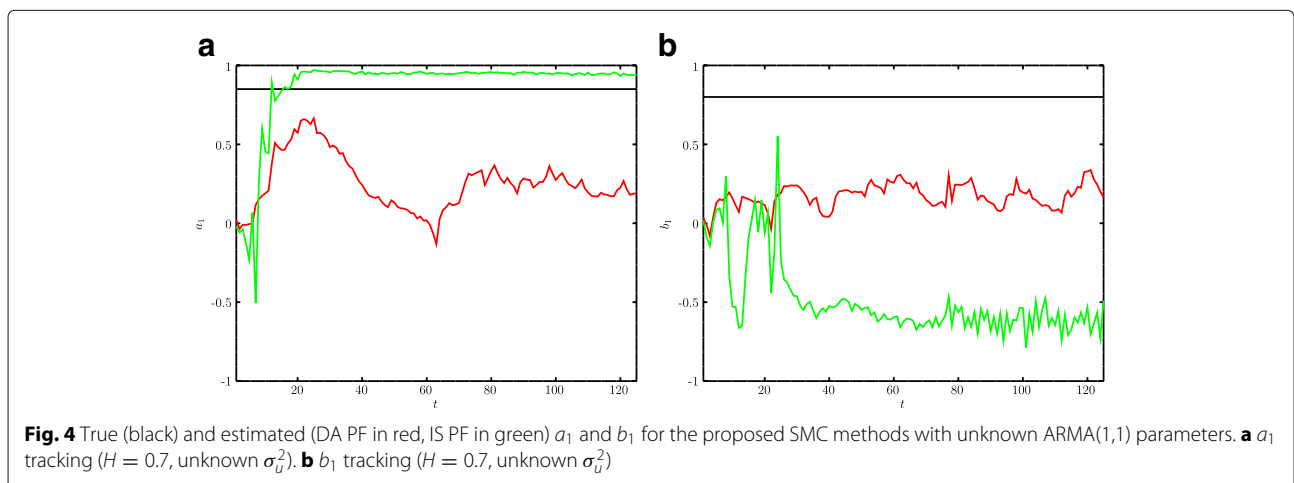
Fig. 4 True (black) and estimated (DA PF in red, IS PF in green) a_1 and b_1 for the proposed SMC methods with unknown ARMA(1,1) parameters. **a** a_1 tracking ($H = 0.7$, unknown σ_u^2). **b** b_1 tracking ($H = 0.7$, unknown σ_u^2)

Fig. 5 True (black) and estimated (DA PF in red, IS PF in green) scale factor for the proposed SMC methods with unknown ARMA(1,1) parameters. **a** ARMA(1,1), $H = 0.5$. **b** ARMA(1,1), $H = 0.7$. **c** ARMA(1,1), $H = 0:9$
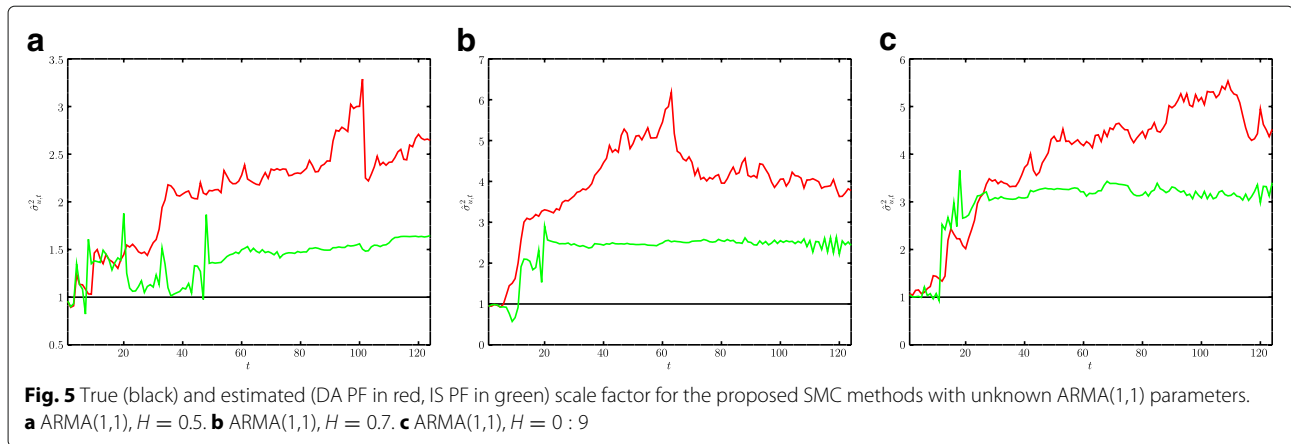
We also observe a slightly better filtering performance of the DA-SMC when compared to the IS-SMC. However, this improved state tracking accuracy comes with a cost, as the estimation of the unknown parameters is worse for the DA-based SMC. This effect is illustrated with some parameter estimation realizations of an ARMA(1,1) process with unknown parameters a_1 and b_1, with known innovation variance σ_u^2 in Fig. 4, and unknown variance σ_u^2 in Fig. 5.

We note that, for both proposed SMC methods, estimation of the AR parameters is more accurate than of the MA parameters. Recall that, due to how these parameters are used in our computations (inversion and multiplication of matrices involved when evaluating the sufficient statistics), identifiability and numerical issues may arise. In the proposed method, particles that are result of numerical instabilities are automatically discarded by their corresponding weights, as they become negligible. In such cases, we observe a reduced effective particle size, but the method is still able to track the state.

Furthermore, the DA-based SMC overestimates the unknown variance (see Fig. 5). Although this might seem irrelevant for the filtering problem, variance overestimation is critical when predicting future instances of the time-series, as the density becomes too wide to be informative.

The poor parameter estimation accuracy for the DA-based SMC can be understood by inspecting the weight computation for each of the proposed alternatives. For the DA-based approach (i.e., Eq. (30)), only state samples are accounted for, while for the IS-based approach as in Eq. (41), both state and parameter samples are taken into account. We further explain this effect with results in Table 3 and interpret them as follows.

When applying IS, one explicitly computes weights based on both the state and parameter samples, while with the DA approach, one hopes that the best state particles are associated with good parameters too (although this is not explicitly accounted for). As a result of the parameter-explicit weight computation in IS-based methods, the number of particles with non-negligible weights is much reduced at every time instant (as one looks for both good state and parameter samples). Consequently, the effective particle size of the IS-based SMC method

Table 3 Efficient particle size M_{eff} of the proposed SMC methods for different ARMA models with fGn

PF type	Effective particle size (M_{eff})					
	Known a, b	Known a, b	Unknown a, b, DA	Unknown $a, b,$, IS	Unknown a, b, DA	Unknown a, b IS
	Known σ_u^2	Unknown σ_u^2	Known σ_u^2	Known σ_u^2	Unknown σ_u^2	Unknown σ_u^2
AR(1), $H = 0.5$	737.19	710.72	753.00	210.89	619.77	173.54
AR(1), $H = 0.7$	712.37	696.36	750.13	194.98	606.86	207.05
AR(1), $H = 0.9$	734.99	724.64	772.48	199.88	628.91	179.13
MA(1), $H = 0.5$	766.83	745.32	784.75	219.95	711.18	193.41
MA(1), $H = 0.7$	753.40	746.07	785.24	209.51	709.84	205.89
MA(1), $H = 0.9$	791.34	779.44	835.39	229.24	741.39	202.39
ARMA(1,1), $H = 0.5$	653.89	662.80	668.61	84.061	530.88	86.080
ARMA(1,1), $H = 0.7$	610.05	605.56	626.75	78.602	466.84	83.643
ARMA(1,1), $H = 0.9$	645.51	638.31	673.09	62.790	504.69	93.404

Urteaga *et al. EURASIP Journal on Advances in Signal Processing*   (2017) 2017:84

Page 14 of 15

is quite low, and thus, the obtained results much more volatile. Averaged effective particle sizes for the proposed SMC methods are provided in Table 3.

6 Conclusions

In this paper, we proposed a set of SMC methods for inference of latent time-series with innovations correlated in time. This is achieved in a unified and consistent manner for different types of models and for scenarios that include known and unknown parameters. We mathematically formulated the problem using the state-space methodology, where the latent time-series was modeled as an $ARMA(p, q)$ driven by innovations correlated in time. The provided compact formulation allows for a Bayesian analysis of the model, which results in the derivation of the key transition density for the proposed SMC methods. Different parameter assumptions were considered and, as shown by the presented extensive results, the SMC methods are able to accurately infer the hidden states. The proposed method is generic in that it addresses diverse memory properties in a coherent manner and is accurate in estimating latent time-series.

Endnote

[1] We use innovations to refer to the stochastic process driving the time-series. Although a reader from the signal-processing community might be more familiar with the term noise, we prefer to use innovations as it is most common in the statistical literature of stochastic processes and time-series analysis.

Authors' information
Iñigo Urteaga received the M.S. degree in telecommunications engineering from the ETSI Bilbao, UPV/EHU, Bilbao, Spain, in 2008, and the Ph.D. degree in Electrical Engineering from Stony Brook University, Stony Brook, NY, USA, in 2016. He is currently a Postdoctoral Research Scientist affiliated with both the Data Science Institute and the Department of Biomedical Informatics, at Columbia University, NY, USA. From 2007 to 2008, he was a Research Assistant in the Department of Computer Science, Colorado School of Mines (USA) and a Research Associate with Tecnalia Research and Innovation (Spain) from 2009 to 2011. He received the 2016 Armstrong Memorial Award for the best graduate student in the Department of Electrical and Computer Engineering at Stony Brook University. His main research interests include statistical signal processing and more specifically, Bayesian statistics and simulation-based Monte Carlo techniques. He focuses on the science of data inference, modeling, and prediction with applications to a wide range of disciplines. Urteaga is a member of the IEEE and serves as a reviewer of several journals and conferences.

Mónica F. Bugallo received her B.S., M.S., and Ph. D. degrees in computer science and engineering from University of A Coruña, Spain. She joined Stony Brook University in 2002 where she is currently a Professor of the Department of Electrical and Computer Engineering, as well as the Faculty Director of the Women In Science and Engineering (WISE) program of the College of Engineering and Applied Sciences. Her research interests are in the field of statistical signal processing, with emphasis on the theory of Monte Carlo methods for complex systems and its application to different disciplines including cancer research, sensor networks, and finance. In addition, she has focused on STEM education and has initiated several successful programs with the purpose of engaging students at all academic stages in the excitement of engineering and research, with focus on underrepresented groups. She has authored and coauthored two book chapters and more than 150 journal papers and refereed conference articles. Bugallo is a senior member of the IEEE, serves as an elected member on the Sensor Array and Multichannel and the Signal Processing Theory and Methods technical committees, and is the current chair of the IEEE Signal Processing Society Education Committee. She has been part of the technical committee and has organized various professional conferences and workshops and has received several prestigious research and education awards including the award for Best Paper in the IEEE Signal Processing Magazine (2007), the IEEE Athanasios Papoulis Award (2011), and the Chair of Excellence by the Universidad Carlos III de Madrid-Banco de Santander (Spain) (2012).

Petar M. Djurić received the B.S. and M.S. degrees in electrical engineering from the University of Belgrade, Belgrade, in 1981 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Rhode Island, Kingston, RI, in 1990. Since 1990, he has been with Stony Brook University in the Department of Electrical and Computer Engineering, where he is a SUNY Distinguished Professor. From 1981 to 1986, he was a Research Associate with the Vinča Institute of Nuclear Sciences, Belgrade. His research interests include the area of signal and information processing with primary interests in the theory of signal modeling, detection, and estimation; Monte Carlo-based methods; signal and information processing over networks; and applications in a wide range of disciplines. He has been invited to lecture at many universities in the USA and overseas. He received the IEEE Signal Processing Magazine Best Paper Award in 2007 and the EURASIP Technical Achievement Award in 2012. In 2008, he was the Chair of Excellence of Universidad Carlos III de Madrid-Banco de Santander. From 2008 to 2009, he was a Distinguished Lecturer of the IEEE Signal Processing Society. He has been on numerous committees of the IEEE Signal Processing Society and of many professional conferences and workshops. He is a Fellow of IEEE and EURASIP and the Editor-in- Chief of the IEEE Transactions on Signal and Information Processing Over Networks.

Author details
[1] Department of Electrical and Computer Engineering, Stony Brook University, 11794 Stony Brook, NY, USA. [2] Department of Applied Physics and Applied Mathematics, Columbia University, 10027 New York, NY, USA.

References
1. PJ Brockwell, RA Davis, *Time Series: Theory and Methods, 2nd edn. Springer Series in Statistics*. (Springer, 1991)
2. J Durbin, SJ Koopman, *Time Series Analysis by State-Space Methods. Oxford Statistical Science Series*. (Oxford University Press, 2001)
3. RH Shumway, DS Stoffer, *Time Series Analysis and Its Applications: With R Examples (Springer Texts in Statistics)*, 3rd edn. (Springer, 2010)

Urteaga *et al. EURASIP Journal on Advances in Signal Processing* (2017) 2017:84

Page 15 of 15

4. P Whittle, *Hypothesis Testing in Time Series Analysis*. (Almquist and Wicksell, 1951)
5. GEP Box, GM Jenkins, *Time Series Analysis: Forecasting and Control. Holden-Day series in time series analysis and digital processing*. (Holden-Day, 1976)
6. HE Hurst, Long-term storage capacity of reservoirs. Trans. Am. Soc. Civil Eng. **116**, 770–808 (1951)
7. RT Baillie, Long memory processes and fractional integration in econometrics. J. Econ. **73**(1), 5–59 (1996)
8. J Beran, *Statistics for Long-Memory Processes. Chapman & Hall/CRC Monographs on Statistics & Applied Probability*. (Taylor & Francis, 1994)
9. W Palma, *Long-Memory Time Series: Theory and Methods. Wiley Series in Probability and Statistics*. (Wiley, 2007)
10. H Liu, E Erdem, J Shi, Comprehensive evaluation of ARMA–GARCH(-M) approaches for modeling the mean and volatility of wind speed. Appl. Energy. **88**(3), 724–732 (2011)
11. DJ Swider, C Weber, Extended ARMA models for estimating price developments on day-ahead electricity markets. Electr. Power Syst. Res. **77**(5–6), 583–593 (2007)
12. R Prado, HF Lopes, Sequential parameter learning and filtering in structured autoregressive state-space models. Stat. Comput. **23**(1), 43–57 (2013)
13. S Degiannakis, C Floros, *Methods of Volatility Estimation and Forecasting*. (Palgrave Macmillan UK, London, 2015), pp. 58–109
14. ST Rachev, JSJ Hsu, BS Bagasheva, FJ Fabozzi, *Bayesian Methods in Finance. Frank J. Fabozzi Series*. (Wiley, 2008)
15. J Durbin, SJ Koopman, *Time Series Analysis by State-Space Methods: Second Edition, 2nd edn. Oxford Statistical Science Series*. (Oxford University Press, 2012)
16. RE Kalman, A new approach to linear filtering and prediction problems. Trans. ASME–J. Basic Eng. **82**(Series D), 35–45 (1960)
17. E Jacquier, NG Polson, PE Rossi, Bayesian analysis of stochastic volatility models. J. Business Econ. Stat. **12**(4), 371–89 (1994)
18. C Broto, E Ruiz, Estimation methods for stochastic volatility models: a survey. J. Econ. Surv., 613–649 (2004)
19. BDO Anderson, JB Moore, *Optimal Filtering. Dover Books on Electrical Engineering*. (Dover Publications, 2012)
20. MS Arulampalam, S Maskell, N Gordon, T Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. Signal Process. IEEE Trans. **50**(2), 174–188 (2002)
21. PM Djurić, JH Kotecha, J Zhang, Y Huang, T Ghirmai, MF Bugallo, J Míguez, Particle filtering. IEEE Signal Process. Mag. **20**(5), 19–38 (2003)
22. A Doucet, N De Freitas, N Gordon, *Sequential Monte Carlo Methods in Practice*. (Springer, 2001)
23. NJ Gordon, DJ Salmond, AFM Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation. Radar Signal Process. IEEE Proc. **140**(2), 107–113 (1993)
24. B Ristic, S Arulampalam, N Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. (A Print, ed.) (Artech House, 2004)
25. PJ van Leeuwen, Particle filtering in geophysical systems. Monthly Weather Rev. **12**(137), 4089–4114 (2009)
26. EL Ionides, C Bretó, AA King, Inference for nonlinear dynamical systems. Proc. Natl. Acad. Sci. **103**(49), 18438–18443 (2006)
27. EL Ionides, KS Fang, RR Isseroff, GF Oster, Stochastic models for cell motion and taxis. J. Math. Biol. **48**(1), 23–37 (2004)
28. D Creal, A survey of Sequential Monte Carlo methods for economics and finance. Econ. Rev. **31**(3), 245–296 (2012)
29. CM Carvalho, MS Johannes, HF Lopes, NG Polson, Particle learning and smoothing. Stat. Sci. **25**(1), 88–106 (2010)
30. N Chopin, PE Jacob, O Papaspiliopoulos, SMCˆ2: an efficient algorithm for sequential analysis of state-space models. ArXiv e-prints (2011). 1101.1528
31. D Crisan, J Míguez, Nested particle filters for online parameter estimation in discrete-time state-space Markov models. ArXiv e-prints (2013). 1308.1883
32. I Urteaga, PM Djurić, in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. Estimation of ARMA state processes by particle filtering, (2014), pp. 8033–8037
33. I Urteaga, PM Djurić, Sequential estimation of hidden ARMA processes by particle filtering—part I. IEEE Trans. Signal Process. **65**, 482–493 (2016)
34. I Urteaga, PM Djurić, Sequential estimation of hidden ARMA processes by particle filtering—part II. IEEE Trans. Signal Process. **65**, 494–504 (2016)
35. I Urteaga, MF Bugallo, PM Djurić, in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015 IEEE 6th International Workshop On*. Filtering of nonlinear time-series coupled by fractional Gaussian processes, (2015)
36. CC Drovandi, JM McGree, AN Pettitt, A sequential monte carlo algorithm to incorporate model uncertainty in bayesian sequential design. J. Comput. Graph. Stat. **23**(1), 3–24 (2014)
37. I Urteaga, MF Bugallo, PM Djurić, in *2016 IEEE Statistical Signal Processing Workshop (SSP)*. Sequential Monte Carlo methods under model uncertainty, (2016), pp. 1–5
38. L Martino, J Read, V Elvira, F Louzada, Cooperative parallel particle filters for online model selection and applications to urban mobility. Digital Signal Process. **60**(Supplement C), 172–185 (2017)
39. JM Bernardo, AFM Smith, *Bayesian Theory. Wiley Series in Probability and Statistics*. (Wiley, 2009)
40. T Li, M Bolić, PM Djurić, Resampling methods for particle filtering: classification, implementation, and strategies. Signal Process. Mag. IEEE. **32**(3), 70–86 (2015)
41. L Martino, V Elvira, F Louzada, Effective sample size for importance sampling based on discrepancy measures. Signal Process. **131**, 386–401 (2017)
42. MH Hayes, *Statistical Digital Signal Processing and Modeling*. (John Wiley & Sons, 1996)
43. J Liu, M West, *Combined Parameter and State Estimation in Simulation Based Filtering*. (Springer, 2001), pp. 197–224. Chapter 10 in "Sequential Monte Carlo Methods in Practice"
44. C Nemeth, P Fearnhead, L Mihaylova, Sequential Monte Carlo methods for state and parameter estimation in abruptly changing environments. Signal Process. IEEE Trans. **62**(5), 1245–1255 (2014)
45. PM Djurić, MF Bugallo, J Míguez, in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*. Density assisted particle filters for state and parameter estimation, vol. 2, (2004), pp. ii701–704
46. J Olsson, O Cappé, R Douc, E Moulines, Sequential Monte Carlo smoothing with application to parameter estimation in non-linear state space models. ArXiv Mathematics e-prints (2006). math/0609514
47. J Olsson, J Westerborn, Efficient particle-based online smoothing in general hidden Markov models: the PaRIS algorithm. ArXiv e-prints (2014). 1412.7550
48. G Agamennoni, EM Nebot, Robust estimation in non-linear state-space models with state-dependent noise. Signal Process. IEEE Trans. **62**(8), 2165–2175 (2014)
49. PM Djurić, M Khan, DE Johnston, Particle filtering of stochastic volatility modeled with leverage. Selected Topics Signal Process. IEEE J. **6**(4), 327–336 (2012)
50. AC Harvey, E Ruiz, N Shephard, Multivariate stochastic variance models. Rev. Econ. Stud. **61**(2), 247–264 (1994)
51. R Bhar, D Lee, Comparing estimation procedures for stochastic volatility models of short-term interest rates (2009)
52. SS Ozturk, J-F Richard, Stochastic volatility and leverage: application to a panel of S&P500 stocks. Finance Res. Lett. **12**, 67–76 (2015)
53. N Shephard, *Stochastic Volatility: Selected Readings. Advanced texts in econometrics*. (Oxford University Press, 2005)
54. SJ Julier, JK Uhlmann, Unscented filtering and nonlinear estimation. Proc. IEEE. **92**(3), 401–422 (2004)
55. RVD Merwe, E Wan, in *Proceedings of the Workshop on Advances in Machine Learning*. Sigma-point Kalman filters for probabilistic inference in dynamic state-space models, (2003)
56. O Zoeter, A Ypma, T Heskes, in *Machine Learning for Signal Processing, 2004. Proceedings of the 2004 14th IEEE Signal Processing Society Workshop*. Improved unscented Kalman smoothing for stock volatility estimation, (2004), pp. 143–152
57. R Cont, in *Fractals in Engineering*, ed. by J Lévy-Véhel, E Lutton. Long range dependence in financial markets (Springer, 2005), pp. 159–179